

Peer-review report of

Bowes, S. M., & Fazio, L. K. (2024). Intellectual humility and misinformation receptivity: A meta-analytic review. *advances.in/psychology*, 2, e940422.
<https://doi.org/10.56296/aip00026>

Round 1

Dear Authors,

Thank you for submitting your intriguing meta-analysis to our special issue in *advances.in/ psychology*. I was fortunate to receive evaluations from two experts in the field. Both have extensive experience with misinformation and conspiracy theories and have conducted meta-analyses previously. The first reviewer has meta-analyses as one of their core areas of expertise. I extend my gratitude to both reviewers for their constructive and detailed feedback. As a standard procedure, I first read and evaluated the manuscript myself before consulting the reviewers' evaluations.

We all recognize many strengths in your work. Your reporting is detailed, and your analyses were conducted meticulously. The paper is generally well-written and efficiently structured. However, there are some areas for improvement to fully realize the potential of the manuscript.

Reviewer 1 suggests presenting 90% prediction intervals (PIs), and I concur. Please ensure your discussion and implications section reflect the evidential value conveyed by the PIs. Additionally, the complexity of many moderation analyses could be mitigated by presenting figures summarizing the results (not just for political orientation). I also agree with the reviewer that moderation analyses with fewer than three studies are less meaningful. While the reviewer proposes a cutoff of three studies, I would suggest even fewer, which will also make the moderation analyses more robust. Reviewer 1 has made many additional comments that should help improve your methods and analyses, and I ask you to respond to each of them in your revision.

Reviewer 2's main concern is the reliance on correlational data and the partly causal language. They recommend moderating the language to avoid implying causal effects and focusing on associations instead of effects/predictors. I agree with this point and ask you to adjust the framing of your study and its novelty accordingly. While we should not disregard the potential causal effects suggested by correlational analyses, we must acknowledge that if a causal effect exists, it is usually much

smaller than the correlations suggest (often less than one-third of the original effect). Additionally, I understand that longitudinal studies were coded correlationally (within wave), but it might be worthwhile to consider whether there are enough studies to conduct a longitudinal (cross-lagged) meta-analysis as an addition to your current analyses. This is not mandatory but could strengthen your argument regarding the consequences of IH. Reviewer 2 has additional points, and I request that you address each of them in your revision.

In addition to the points outlined above, I have the following comments. Some of these may overlap partly with those of the reviewers.

METHODS AND REPORTING

- As Reviewer 2 noted, it was unclear why terms such as “conspiracy theor*” were not included as search terms. Including common scales like “conspiracy mentality” could also have been beneficial. While it is uncertain how many extra studies this would yield, this limitation should at least be discussed. The same applies to the selection of databases (Reviewer 1).
- More information on how the Google search was conducted would be beneficial.
- As Reviewer 1 suggested, please include reliability estimates between coders.
- While your attempt to provide as much information as possible through Table 1 is commendable, it is a bit overwhelming. Consider moving some information, such as the countries, to the Supplementary Online Material (SOM). Additionally, the column for the number of significant effects may not be very meaningful.
- It is unclear how “percentage of Democrats” was coded for non-US studies. Please confirm and explain this in the text (my assumption it was only coded for studies from the US).
- Effect sizes in visualizations such as Figure 2, 4, and 6 should be sorted by size (maybe except for the overall effect that always can be presented at the end). The legend on the right is unnecessary, and different colors (which are not needed) may be challenging for individuals with visual impairments. For aesthetics, please provide the effect sizes on top of the line/point estimates in small letters rather than next to them.
- In the results, you write, “the studies were predominately comprised of female ($M = 54.93\%$, $SD = 14.27$),” but wouldn’t it be more accurate to describe this as close to gender-balanced?
- Please present Table 2 in landscape format to maintain margins.
- When reporting effects in the text, avoid repeating statistics already presented in the table.

- Given the complexity of the moderation analyses, it would be helpful to include effect size plots for these as well.

MINOR COMMENTS

- I believe you mean “inconsistent” on p. 16?
- Please report the R version used and cite the software itself (not just R Studio).
- On p. 18 and other relevant places, please ensure to write out acronyms when first introduced.
- There are some punctuation issues (e.g., missing period in “et al.”).
- When first introducing “Gignac and Szodorai’s (2016) effect size guidelines,” please specify the cutoffs briefly. This relates to Reviewer 2’s comment. There is a lot of controversy regarding Cohen’s cutoffs. As long as it is well-argued for, I think adopting more modern conventions are fine.
- Please revise this sentence that seems to be somewhat circular: “Mirroring belief in fake news, the relationship between IH and less willingness to share fake news was significantly weaker for COVID-19 fake news ($b = -.05$, $p = .280$) compared to other fake news measures, all of which were weak, positive, and significant predictors of less willingness to share fake news (bs ranged from .09 to .14, $ps < .05$).”
- Ensure to provide DOIs in URL format for all relevant references in line with APA7.

Thank you for submitting this very promising research to our special issue. While we ask you to address a relatively large number of requests, they should not be too work-intensive and are relatively easily addressable. This will make your analyses and interpretations more efficient and robust, enhancing the impact of your paper on the field. I look forward to receiving your revision, accompanied by a letter outlining all changes, including excerpts, within a month. Please reach out if you need more time.

Best regards,

Jonas R. Kunst

Reviewer 1:

The proposed meta-analysis examines the relation between intellectual humility and misinformation beliefs, intentions, and behaviors, with results suggesting that intellectual humility might indeed have some effects. While overall I consider this

topic interesting and the paper worth publication, I have several comments that I hope will help the authors improve their report.

Introduction

The introduction is well-written and convincing. The only remark I have here is that the authors try to argue for their study by recurring to novelty claims, which strike me as unconvincing and rather unnecessary, since the paper at hand is a meta-analysis of already existing (i.e., not that novel) research:

“There is scant research on whether and how certain individual differences characteristics may protect against misinformation” – I am not sure if that is true. Stasielowicz (2022) and Goreis & Voracek (2019) published entire meta-analyses on the link between conspiracy beliefs – a form of misinformation – and individual differences.

“In comparison, less research has focused on the latter, that is identifying the factors that reduce misinformation receptivity” – again, I am not sure this is true. I would say it all depends on how you look at the tested variables: a lack of education increases it/having education decreases it.

I would suggest removing such arguments, and instead simply stating that there is now enough volume of research on IH to conduct a meta-analysis.

Methods & Results

1. Some important elements are not reported. For instance, abstract screening and coding should be done (at least in part) by two coders, and interrater agreement should be provided. I suggest that the authors check the APA MARS guidelines and ensure that all the listed aspects are reported.

2. I would generally not consider PsychInfo and Google as a sufficient database coverage for literature search. Especially for conspiracy theories, many publications are not indexed in PsychInfo but are listed in PubMed/Medline. Please provide a justification for this database choice. Also, please clarify in the paper if your 30-paper rule applies to Scholar specifically.

3. The convention in meta-analysis is that k refers to the number of studies, which is the same as the number of independent samples, whereas the number of reports is distinct from it. I admit that the literature can be chaotic in this regard, but I would strongly advise against using the term “study” when referring to a report; it is confusing, and a reader less familiar with meta-analysis may struggle to understand why your “study” number k is not part of your model specification.

4. Data analysis plan: The authors state that correlations were weighted by their sampling error, but the code makes it clear that they used (correctly) the standard inverse variance weighting. It is not the same, and I urge the authors to correct this sentence.

5. I was confused by the t-test follow up approach to categorical moderators. Was this done within the meta-regression model by removing one category of the moderator? If so, please explain this approach more clearly and provide references, or re-run the meta-regressions using the standard approach (i.e., include all categories of the predictor and alter the baseline category to compare each pair of categories; note that this is likely to change the results).

6. “Subgroups were only included in these moderation analyses if there were at least two effect sizes present for that subgroup.” – this seems severely underpowered, please reconsider. At the very least, I would suggest three effects from different samples.

7. Heterogeneity: I was happy to see that the authors included a good deal of heterogeneity information. However, they opted for reporting and interpreting I^2 and H^2 . I am not convinced that H^2 adds anything over I^2 (which, by the way, is also – theoretically – independent from k – as in number of samples, not reports; empirically, neither of them are). These statistics are both ratios and neither of them is particularly informative, but I am not opposed to reporting either of them, as long as the authors also report prediction intervals (90% PI is fine) and interpret these. This will help put the results in perspective. For example, a quick look at Table 2 suggests that only the results relative to Intentions: Investigate Misinfo and Behaviors are likely to be actually meaningful according to the PIs (but underpowered, thus at risk of Type 1 error).

8. Side note: the figures reported as forest plots are generally known as caterpillar plots.

Discussion

The discussion is extensive, and the main issue I see is that it is, at times, overly optimistic in ways unjustified by the results. Considering the extreme heterogeneity of most models, I would strongly advise against statements such as “IH was a small to moderate but consistent predictor” because the overall results do not point to consistency at all. Instead, I suggest (slightly) re-writing the discussion by factoring in the information provided by the prediction intervals, and generally including more of the uncertainty around the results that come from both the heterogeneity and the low power for some subgroup analyses.

Reviewer 2:

I thank the authors and editor for the opportunity to review this submission. It presents a meta-analysis of the relationship between intellectual humility and misinformation receptivity. The average relationships estimated in the meta-analysis varied somewhat across various moderating factors and forms of misinformation receptivity, but were generally fairly small.

Before continuing I should note that I am not an expert in meta-analyses (I have been a co-author on just one and never led one myself). I trust there will be at least one other reviewer with more technical and statistical expertise in this specialist area.

I think the topic of misinformation receptivity is an important one. I thought the study was very clearly and transparently reported, and thorough to a fault. I also appreciated the organised open data and materials. It was good to see a contemporary method for tackling publication bias (PET-PEESE etc.). To the limited extent of my knowledge I could find little in the way of methodological problems.

I nevertheless do have one relatively major issue to raise that might require fairly substantial revisions. Specifically, I notice that the rationale for the study seems to rely heavily on the idea that intellectual humility might have a causal effect on misinformation receptivity. Some quotes from the abstract and introduction that demonstrate this are below.

- “factors that **increase** misinformation receptivity”
- “What **makes** some people more receptive to misinformation and what makes others more resilient against it?”
- “what **fosters** resilience against misinformation”
- “IH may **buffer** against misinformation”
- “IH may **protect** against the harmful effects of misinformation”
- “an intriguing individual differences construct that may **reduce** receptivity to misinformation: intellectual humility”, etc.

These are all implicitly theorising about causal effects. The problem here is that the body of research reviewed in this paper is observational, and the ‘effect sizes’ are just correlations. Consequently they tell us next to nothing about the effect of IH on misinformation receptivity; the correlations observed in this study could very well be due to an effect in the opposite direction, or any of a myriad of environmental or genetic third variables. I’m thus left wondering why we need to know about the average size of these correlations; what practical or theoretical gap in knowledge can this meta-analysis fill? I.e., why does this study *matter*? I’d suggest there needs to be substantial revision to the introduction so that it’s clear what gap in knowledge this study aims to fill even if it can’t inform us about causal effects.

Relatedly, the discussion and abstract also include quite a number of implicit causal inferences. Some examples of problems in this regard are below, including several

examples of action recommendations that would only be warranted if the study provided evidence of causal effects.

- “These results suggest that a blend of IH features, including metacognitive, relational, and emotional IH features, might **confer** more resilience against misinformation relative to metacognitive IH features in isolation”
- “By intervening on intellectual humility, it may be possible to **reduce** belief in misinformation and promote intentions to and behavioral engagement with practices that move people away from misinformation.”
- “Intellectual humility may be an especially effective **intervention** for misinformation receptivity”
- “By intervening on IH, it may be possible to not only help people interact with new information in a more accurate way but also to build domain-general abilities and motives that help people hold their preexisting beliefs more lightly”
- “2024). IH may help **move** the needle even in those with strong views.

I’d suggest the discussion and abstract should be revised to either avoid implicit causal inferences, or to make it clearer that these suggestions are speculative rather than warranted by the findings, along with the assumptions they’re based on.

MINOR POINTS

1. Relating to my points above about causality, I’d encourage the authors to refer to the statistics in ways that minimise the possibility of readers misinterpreting correlations for causal effects. E.g, “relationship” or “correlation” will often be a better term than “effect”; “correlate” or “relate” is clearer than “predict”. I recognise it will be hard to do this completely given the degree to which the term “effect size” permeates the literature on meta-analysis even for correlational studies, but there’s still definitely room for some refinement here.
2. I know that this journal has a generous 10k limit for review papers, but I did find the manuscript very long relative to its content. Part of the issue here was the meta-analysis, while admirably thorough, was not particularly focused: There were many subtypes of misinformation receptivity examined, along with many possible moderators. The fact that the authors end up reporting moderation results only for the significant findings is a bit of a sign that too many analyses may have been on the table. I’m very hesitant to say the authors should change this now, but personally I’d have preferred a manuscript more tightly focused on specific research questions and hypotheses of particular theoretical or practical importance.
3. P. 9, I didn’t quite follow how sources of heterogeneity in the meta-analysis would be “boundary conditions”.

4. P. 10, “We also searched through the references sections of two papers (Bowes & Tasimi, 2022; Porter et al., 2022).” Why these two papers in specific, and only them?
5. P. 10, re. the search terms (“conspiracy belief” OR “conspiratorial ideation” OR “conspiracist belief” OR “conspiracist ideation”): Why not just search for “consp*”? Researchers doing work on conspiracy theories don’t always use one of these terms (c.f., “belief in conspiracy theories” etc.)
6. P. 10, “We implemented a stop rule in our search such that if 30 studies in a row”. 30 in a row when ordered in what way?
7. P. 11, “Pearson’s r coefficients were extracted from each of the studies”. Were these always zero-order coefficients, or were any partial correlations?
8. P. 23 (and elsewhere): $r = .29$ would not usually be described as a “strong” correlation; this means each variable explains less than 10% of the variance in the other. According to Cohen’s classic guidelines, this is a “medium” correlation. Please update the phrasing throughout to avoid inadvertently exaggerating the sizes of the relationships found.
9. Please try to ensure the metric/statistic is clear in all figures (e.g., in Figure 4 are “meta-analytic effects” in the X axis correlations?
10. Given my comments about causality above this is something of a moot point, but even the relationships estimated in this study were unbiased estimates of the effect of intellectual humility on misinformation receptivity, I’m still not sure that they’d warrant much enthusiasm about intellectual humility as a target for intervention. Intellectual humility may not be trivial to change with more than a small effect size, and if you multiple that by a (possible) small effect of IH on misinformation receptivity, you would end up with a very small indirect effect.

Round 2

Dear Authors,

Thank you for your responsiveness to the previous round of reviews. We were all impressed by how thoroughly and effectively you addressed the highlighted issues. As a result, this already strong paper has become even more robust and focused. I believe we are getting significantly closer to reaching the level needed to publish the paper. However, some attention to a few points may still be necessary.

I agree with Reviewer 1 that it is not optimal (and unusual) that the reliability of data coding was not assessed. I therefore strongly encourage you to have a portion of the data (e.g., 10%) coded by a second coder (the clearly optimal solution) or by the same coder at a different time (sub-optimal, but perhaps acceptable), and then to

calculate and report the reliability. The reason why the first approach is strongly favored is that it provides insight into the reproducibility of the coding process across researchers.

Please also discuss the 30-paper cutoff selection rule as a limitation. Disregard the third point raised by Reviewer 1, as they reached out to me directly after submitting the review, noting that their comment did not apply after having revisited your code.

I also agree with Reviewer 2 that the "Causality" section needs to be significantly revised. I suggest removing the heading (so that it becomes a part of the limitations section) and condensing the content. Please clarify that what you provide is not a test of causality. I understand this is not what you intended to convey, but some parts could be interpreted as such. For instance, can you rephrase the sentence "One of the goals of this meta-analysis was to determine if a causal link between IH and misinformation receptivity was feasible" to "One of the goals of this meta-analysis was to determine if the correlational precondition of a causal link between IH and misinformation receptivity was supported"? Here, you can discuss that you tested only one of the three criteria for causal inference and explore how the other two criteria (temporal precedence and non-spuriousness) can be addressed in future research (something you already partly do).

In addition to the points highlighted by the reviewers, please also address the following:

1. You somewhat vary in your use of the acronym "IH," primarily in the introduction. I suggest using it consistently throughout the paper after it has been defined, and consider repeating the definition of the acronym once more at the beginning of the discussion section.
2. The figures are very informative. Thank you for adding them. Could you add jittered dots to them that show the individual effects? This maximizes the visual insights into the data.
3. Please ensure that all figures are added to the document as lossless SVG files. This will help maintain the highest possible image quality and crispness during production.
4. Can you add the overall effect sizes for beliefs, intentions, and behavior to the abstract? This will help communicate the striking finding that the link to behavior is strongest.
5. In terms of interpretations of PIs, you write "Most of the 90% prediction intervals included zero, indicating that some individual estimates in future studies will not be significant even when adopting the same methodological designs." (you write something similar another place). My understanding is that this PI range also suggest that there sometimes may be true effects in the opposite direction (relevant

also to Gollwitzer et al., 2024, who you cite). If my understanding is correct, please adjust the respective sentences; otherwise provide a rebuttal.

6. A very minor issue: the letter "s" in the word "statistics" is italicized here: "Thus, we also examined the H^2 and I^2 statistics".

7. Please read through this citation analysis and implement all changes as there are some inconsistencies regarding the citations and references:
https://drive.google.com/file/d/1w4Mm80PF1eYHQz0CVPEqvYKw_1CKXpCS/view?usp=sharing

Thank you again for submitting this intriguing piece. These comments are intended to make an already impressive work even stronger. I plan to make my final decision without consulting the reviewers in the next round. To facilitate a prompt process, I would prefer to receive your revised manuscript by September 30. Please do not hesitate to reach out if you require more time.

Best,

Jonas R. Kunst

Reviewer 1:

Thank you for your thorough revisions to the manuscript, and your thoughtful and constructive responses to our reviews. I think you've sensibly addressed my earlier points in your revisions. At this stage I just have a couple of minor remaining comments.

1. I think that having a Causality subsection in the Discussion is a nice idea. The first paragraph of this is a good start. The second and third paragraphs are a bit shakier, though. These paragraphs almost seem to be assuming that the paper shows evidence of small causal effects, and then considering the practical implications that small effects might have. But we don't actually know that the true effects are small. They could be zero; they could be negative; they could even be large*. We just don't know. You could present this information in a way where it's clearer that you are evaluating the consequences of the *assumption* that small effects are present, but that would then run into a second problem... I.e., while you argue that small effects can compound over time, the relationships estimated in this study come from individual differences research. Most of the people in the samples with higher levels of IH have probably *already had* years or decades of high levels of IH, and plenty of time for IH to exert an effect on misinformation receptivity (if it does have such an effect). Yet we still see small correlations... So unfortunately while I find these paragraphs interesting I do think they're best deleted, and I think the subsection would be fine without them.

2. "One such risky test would be to examine IH as an intervention for misinformation receptivity." It's perhaps not super clear here what idea you'd test by doing this. Are you just meaning to test the idea that IH has a causal effect on misinformation? (If

so, yep, this does seem like the most sensible way to test that idea!)

Those two small points aside, I'm impressed by this paper, and I really appreciate the work you've done on the revisions.

*While relatively unlikely, this could happen if there exists some third variable that affects IH negatively and misinformation receptivity positively, or vice versa. In that scenario, the correlations could be *downwardly* biased estimates of the causal effects.

Reviewer 2:

Overall, I was quite satisfied with the way the authors revised the paper and addressed my comments. I specially appreciate that they have now included one more database. However, I was not fully convinced regarding three points.

1. The authors have now revealed that there was no double-coding. While I sympathize with their argument that one expert in the field is better than two coders with inadequate levels of training, I still consider second coding as a crucial instrument for quality and reliability assurance in meta-analysis. From my own experience, being an expert hardly protects against human error, hence single coding is just not enough. I strongly recommend that the authors arrange for second coding of a reasonable portion of the data. Note that it doesn't have to be done by a new coder. Since coding was completed a while ago, I would accept if the first author recoded (independently from their first coding) at least the key numbers, such as the effect size and the scores on moderator variables.

2. The authors have revealed that their 30-paper stop rule was applied to all databases. This is very problematic, in my view, as this rule assumes that the sorting by relevance in PsychInfo and PubMed can be trusted. That is just not the case. These databases rely merely on the frequency of keywords, not on full content. LLMs for literature review, such as Rryan or ASReview, have been developed for this exact reason; these models not only use keywords, but also learn how to order papers by relevance based on the full semantic network within the title and abstract. For such models, the use of a stop rule makes sense, but unfortunately not in PsychInfo or PubMed. Google is an exception here because it's merely a search engine and the recommendation is to use it as a complementary search avenue anyway.

I will leave to the Editor the recommendation as to what to do in this situation. In my view, the best course of action would be to redo the title and abstract screening using a program like ASReview, but I understand that at this stage, this is not

practical. The absolute minimum would be to acknowledge this as a limitation and provide some arguments to defend this rule. I personally cannot think of any, but I'm sure the authors had a rationale for this decision.

3. The description of the t-tests is still not sufficient, in my view. I have checked the R code provided by the authors on OSF and some of the R code for Bowes et al., 2023, and in both cases a standard t-test from basic R, meant for primary data seems to have been performed. The problem is that in a meta-regression, all primary effects are weighted by v_i and τ^2 ; a standard t-test does not do that, hence the result are invalid. Please clarify if anything was done with the effects within sub-samples to ensure that their weights were correct and that the t-test was adequate for meta-analytical data.